UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/730,767	12/08/2003	Shingo Kiuchi	9333-361	3437

74989 7590 05/14/2009

ALPINE/BHGL
P.O. Box 10395
Chicago, IL 60610

EXAMINER	
WOZNIAK, JAMES S	

ART UNIT	PAPER NUMBER
2626	

MAIL DATE	DELIVERY MODE
05/14/2009	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

UNITED STATES PATENT AND TRADEMARK OFFICE

Commissioner for Patents
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450
www.uspto.gov

BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

Application Number: 10/730,767
Filing Date: December 08, 2003
Appellant(s): KIUCHI ET AL.

Mr. James P. Naughton
For Appellant

EXAMINER'S ANSWER

This is in response to the appeal brief filed 10/14/2008 appealing from the Office action mailed 5/27/2008.

(1) Real Party in Interest

A statement identifying by name the real party in interest is contained in the brief.

(2) Related Appeals and Interferences

The examiner is not aware of any related appeals, interferences, or judicial proceedings which will directly affect or be directly affected by or have a bearing on the Board's decision in the pending appeal.

(3) Status of Claims

The statement of the status of claims contained in the brief is correct.

(4) Status of Amendments After Final

The appellant's statement of the status of amendments after final rejection contained in the brief is correct.

(5) Summary of Claimed Subject Matter

The summary of claimed subject matter contained in the brief is correct.

(6) Grounds of Rejection to be Reviewed on Appeal

The appellant's statement of the grounds of rejection to be reviewed on appeal is correct.

(7) Claims Appendix

The copy of the appealed claims contained in the Appendix to the brief is correct.

(8) Evidence Relied Upon

4,885,791	FUJII et al	12-1989
6,324,509	BI et al	11-2001
6,975,993	KEILLER et al	12-2005

(9) Grounds of Rejection

The following ground(s) of rejection are applicable to the appealed claims:

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

Claims 1, 3-8, 12-15, 17-18, and 20 are rejected under 35 U.S.C. 103(a) as being unpatentable over Fujii et al *(U.S. Patent: 4,885,791)* in view of Keiller *(U.S. Patent: 6,975,993)* and further in view of Bi et al *(U.S. Patent: 6,324,509)*.

With respect to **Claims 1 and 8**, Fujii discloses:

Identifying a start position of a speech region of speech data for which speech recognition is to be performed and Generating, from speech data for which speech recognition is to be performed, a plurality of pieces of speech data whose start positions of non-speech regions differ *(generating plural possible speech periods having different starting boundaries including varying amounts of unvoiced sounds and noise, Col. 8, Lines 11-49)*; and

Performing speech recognition using each of said pieces of speech data to obtain a plurality of recognized results *(performing pattern matching using the plural possible speech segments, Col. 8, Lines 11-49)*.

Although Fujii discloses the generation of a plurality of possible speech segments for recognition, which each have different starting boundaries including varying amounts of unvoiced sounds and noise and performing speech recognition using those segments, Fujii does not teach providing a speech recognition result using a metric based on the identified most numerous recognized result from among a plurality of obtained recognized results. Keiller, however, recites a plurality of recognition engines utilizing such a metric *(most commonly occurring word or words as recognition result, Col. 21, Lines 1-11)*.

Fujii and Keiller are analogous art because they are from a similar field of endeavor in speech recognition systems. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to modify the teachings of Fujii with the recognition means utilizing the aforementioned scoring metric as taught by Keiller in order to provide a more efficient multi-engine speech recognizer capable of providing a most likely result *(Keiller, Col. 2, Lines 4-8; and Col. 21, Lines 1-11)*.

Although Fujii further discloses predetermined speech period offset times to include varying amounts of non-speech data *(Col. 10, Line 67- Col. 11, Line 20)*, Fujii does not specifically suggest that this plurality of segments is obtained by shifting backwards. Such a backward shift for determining a starting point *(or multiple starting points in the case of Fujii)* of a speech data segment is well known in the speech processing art however, as is evidenced by the Bi reference *(Col. 5, Lines 13-30)*.

Fujii, Keiller, and Bi are analogous art because they are from a similar field of endeavor in speech recognition systems. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to modify the teachings of Fujii in view of Keiller with the concept of backwards searching (shifting) taught by Bi in order to provide a well-known means of achieving the multiple speech data periods in Fujii that can be easily implemented in a real-time processor *(Bi, Col. 5, Lines 24-30)*.

With respect to **Claim 3**, Bi further shows a speech segment endpointer, which determines a speech starting point, as part of a speech recognizer *(Fig. 1, Element 22)*.

With respect to **Claims 4, 12, and 20**, Bi discloses the means for determining a speech segment starting point in a speech recognizer, as applied to claim 3, while Fujii discloses that the period of this input segment can be varied to account for an uncertain amount of non-speech data, as applied to Claim 1. Since the period of the speech data is varied only based on an uncertain amount of non-speech data, the speech region would be the same for the plurality of generated segments in Fujii, and thus, identical to the first speech data starting point determined by the endpointer taught by Bi.

With respect to **Claims 5, 13, and 17**, Fujii further discloses an A/D conversion of an input speech signal at a predetermined sampling frequency *(Col. 8, Lines 14-16)*, while Bi discloses an circular buffer that stores a sequence of speech data frames in order *(Col. 5, Lines 13-30)*. Bi also discloses changing a buffer reading position to determine a speech data starting point, as applied to Claim 2.

With respect to **Claims 6, 14, and 18**, Fujii discloses that individual speech samples are obtained at a rate of 8kHz *(Col. 8, Lines 14-16)*.

With respect to **Claim 7**, Keiller discloses the multi-engine speech recognizer as applied to Claim 1.

Claim 15 contains subject matter similar to Claims 7 and 8, and thus, is rejected for the same reasons.

(10) Response to Argument

With respect to independent claims 1, 8, and 15, the appellants argue that: the 35 U.S.C. 103(a) rejection is based upon mere conclusive statements, Fujii et al *(U.S. Patent: 4,885,791)* does not teach determining the actual starting point of speech and obtaining different speech periods by shifting backwards, Bi et al *(U.S. Patent: 6,324,509)* does not describe sequentially shifting backwards to obtain a plurality of starting points and determining a beginning point that is the start of a speech period *(Appeal Brief, Pages 7-10)*. The appellants also further make some newly presented additional arguments directed towards the dependent claims *(Appeal Brief, Pages 10-11)*. These arguments will be specifically addressed by the examiner below.

With respect to independent claims 1, 8, and 15, the appellants first argue that Fujii, the primary reference, teaches that error in detection of a speech period are caused by noise and address this problem by not determining the actual starting point of speech, but multiple proposed speech periods *(Appeal Brief, Page 8)*. The appellants continue to argue that in their invention, an actual starting point is identified and then varying degrees of a preceding non-speech region are added thereto *(Appeal Brief, Page 8)*.

In response, the examiner first notes that the appellants' own invention is also not concerned with determining an *actual* starting point of speech. Instead, the appellants' invention involves determining a plurality of starting points in a supposed non-speech region in order to cope with a high noise level in a recognition environment *(Specification, Page 3, Paragraph 0010)*. In other words, both the present invention and Fujii, add a certain amount of a supposed non-speech region to a speech region in order to cope with high-noise speech recognition environments. In Fujii, it is specifically noted that "in order to avoid ambiguity...due to...noise introduction, plural possible speech periods are extracted" (Col. 8, Lines 29-49).

The examiner secondly notes that the appellants' *claimed* invention is silent on determining an *actual* starting point of speech. For instance, in Claim 1, Line 3; Claim 8, Line 2; and Claim 15, Line 3, the claims specifically merely set forth "identifying a start position". Thus, the claims do not require that an *actual* start position be found as is argued by the appellants, only that a general "start position" is found from which non-speech regions can be added by sequentially shifting backwards therefrom. In response to this argument that the references fail to show certain features of applicant's invention, it is noted that the features upon

which applicant relies *(i.e., the identification of an "actual" starting point)* are not recited in the rejected claim(s). Although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993). With respect to the claimed invention then, Fujii teaches that a most likely speech period is determined though a power level measurement and multiple candidate speech periods having varying start positions are acquired therefrom by the addition of unvoiced speech or noise *(Col. 8, Lines 11-49)*. Bi also teaches the same use of a signal level in determining a likely speech start position (PRE_Start) from which multiple candidate starting points in a unvoiced/noisy region can be considered/generated by sequentially shifting backwards from that PRE_Start position *(pointer position backwards offset from a PRE_Start position in a buffer containing a segment of speech data is determined to consider multiple candidate start positions, Col. 6, Line 55- Col. 7, Line 24; and Fig. 3)*. Thus, since the combination of Fujii and Bi teaches "identifying a starting position" as is described above and the claim does not require determine an "*actual* starting point of speech", these arguments have been fully considered, but are not convincing.

On Pages 8-9 of the Appeal Brief, the appellants note that Fujii is silent no how specifically to determine the starting points of the multiple candidate speech periods that are submitted to a speech recognizer *(which was also noted by the examiner, See Final Office Action from 5/27/2008, Page 6, Last Paragraph)*. In response, the examiner notes that it is instead the Bi reference that is relied upon to provide this teaching. Further, in response to this argument against the references individually, one cannot show nonobviousness by attacking references individually where the rejections are based on combinations of references. See *In re Keller*, 642

F.2d 413, 208 USPQ 871 (CCPA 1981); *In re Merck & Co.,* 800 F.2d 1091, 231 USPQ 375 (Fed. Cir. 1986). Thus, this argument has been fully considered, but is not convincing.

Next, the appellants address the Bi reference. The appellants first allege that Bi fails to teach determining multiple speech periods having different starting points because Bi's process only ends by determining a single set of start and stop points *(Appeal Brief, Page 9).* The appellants continue to argue that the passage pointed to in Bi merely notes that signal data is stored in a buffer so that a processor can look back a certain number of speech frames *(appeal Brief, Page 9).*

In response, the examiner notes that the important aspect of Bi is that while Bi does attempt to determine a probable starting point, as is argued by the appellants, Bi's process involves the generation and consideration of *several* start position candidates. Bi's generation of multiple start points can be best understood by looking to the illustration shown in Fig. 3 in light of the explanation provided in Col. 6, Line 24- Col. 7, Line 24. More specifically, Bi describes that an input audio signal is accepted form a user and stored in a speech data buffer that enables a "look back" *(Col. 5, Lines 12-31).* Analysis then begins on this stored signal by first marking an index position in the buffer as being a likely starting position using signal levels *(i.e., SNR) (Col. 5, Line 64- Col. 6, Line 22).* The name of this marker is PRE_Start *(see Fig. 3)* and corresponds to the appellants' start position from which a shifting back is to occur. Also worth pointing out in this figure is that the system of Bi determines the PRE_Start position as likely being the start of speech analysis because the speech signal waveform has particularly high levels here, whereas beforehand it is relatively weak and might correspond to speech or noise. Continuing with the

process of Bi, the next step involves *decrementing* the buffer pointer back in time from the PRE_Start position *(Col. 6, Line 55- Col. 7, Line 24)*. This process of decrementing or shifting backwards in a buffer continues until multiple candidate starting positions have been analyzed and a definitive start endpoint has been identified *(Col. 7, Lines 20-31)*. In Fig. 2, Bi additionally shows in Element 136 that a start position buffer index *(I)*, which was set to PRE_Start in Element 122 is sequentially decremented or shifted backwards in the buffer *(I--)* to generate multiple *candidate* speech periods for analysis. In Fig. 3 it is shown that Bi's system sequentially shifts backwards in time form a PRE_Start position, considering multiple candidate periods with different starting points until a most likely extended period *(START)* is found. Thus, the main concept taught by Bi that is relied upon in the claim rejection is not the eventual acquiring of a most likely extended period which includes the stronger speech period defined by a PRE_Start position and an additional weaker unvoiced or noisy section added by the backward-shifting analysis *(as is argued by the appellants)*, but the generating and considering *multiple* candidate periods in the process of that analysis by buffer pointer decrementing. In Fujii, generating multiple candidate speech periods to be applied to a recognizer is important due to ambiguous speech periods resulting from unvoiced speech or noise *(Col. 8, Lines 50-54)* and Bi provides a specific way for Fujii to generate these *candidates* by sequentially shifting backwards in an audio buffer. Applying this strategy to Fujii provides the benefit of providing a specific technique to efficiently generate these candidate periods using a real-time processor *(Col. 5, Lines 24-30)*, while also ensuring that no weak speech segments are missed *(Col. 9, Lines 29-30)* due to noise or unvoiced speech *(as is considered by Fujii)* as a result of this step-by-step look back process. Thus, since Bi generates *several candidate* speech periods for analysis and Bi

employs the use of a buffer specifically to shift backwards for a specific benefit, these arguments has been fully considered, but are not convincing.

The appellants next argue that Bi does not "describe sequentially shifting backwards to obtain a plurality of starting points" because the PRE_START point is only an interim calculation point in the process of eventually determining an actual starting point *(Appeal Brief, Pages 9-10)*. In response, the examiner notes that the appellants' invention is not directed to determining an *actual* starting point *(see corresponding response above)*, it only involves finding a "start position" from which shifting analysis/calculations may begin. That is exactly the function of Bi's PRE_Start buffer index *(i.e., the appellants' "start position" is an interim point from which multiple candidates are generated by shifting backwards)*. As described above, the system in Bi sets a starting analysis index to PRE_Start which marks the beginning of a section where speech is likely to begin *(See Fig. 3)*. From there, Bi generates multiple extended candidate periods to take into account weak or noisy speech *(Col. 9, Lines 29-30)* by sequentially decrementing this index position backwards in time *(Col. 6, Line 55- Col. 7, Line 24; and Fig. 2, Element 136)*. Thus, since the PRE_Start position in Bi corresponds to the appellants claimed "start position" *(no "actual" starting position is claimed)*, Bi generates multiple candidate positions by decrementing or shifting backwards in a buffer from the PRE_Start position, and Fujii expresses the desirability of applying multiple speech periods to a speech recognizer *(Col. 8, Lines 50-54)*, these arguments have been fully considered, but are not convincing.

The appellants' also reiterate their argument that Bi only determines one actual starting point *(Appeal Brief, Page 9)*. In response, the examiner notes that in the process of determining a likely starting point, Bi more importantly generates several candidate starting positions by

decrementing a buffer index to cope with weak or noisy speech portions *(see above)*. Since Fujii notes the importance of considering multiple candidate periods *(Col. 8, Lines 50-54)* and Bi provides a beneficial means by which to achieve these *candidate* periods *(efficiently generating these candidate periods using a real-time processor, Col. 5, Lines 24-30, while also ensuring that no weak speech segments are missed, Col. 9, Lines 29-30)*, this argument has been fully considered, but is not convincing.

The appellants finally argue that Fujii and Bi cannot be combined because they are incompatible and that the examiner has relied upon hindsight reasoning in making the corresponding 35 U.S.C. 103(a) rejection *(Appeal Brief, Page 10)*. In response, the examiner notes that, as described above, Bi considers multiple *candidate* starting positions in a noisy or weak speech region by shifting back *(i.e., decrementing a buffer pointer to sequentially shift back in time)* from a starting point *(PRE_Start)* and Fujii expressly notes the importance of considering multiple candidate periods *(Col. 8, Lines 50-54)*. Also, in response to applicant's argument that the examiner's conclusion of obviousness is based upon improper hindsight reasoning, it must be recognized that any judgment on obviousness is in a sense necessarily a reconstruction based upon hindsight reasoning. But so long as it takes into account only knowledge which was within the level of ordinary skill at the time the claimed invention was made, and does not include knowledge gleaned only from the applicant's disclosure, such a reconstruction is proper. See *In re McLaughlin*, 443 F.2d 1392, 170 USPQ 209 (CCPA 1971). Thus, for at least the preceding reasons, these arguments have been fully considered, but are not convincing.

With respect to dependent Claim 3, the appellants newly argue that Bi fails to teach that that the "start position of the speech region is provided by a speech recognition engine which performs the speech recognition" because Bi's endpoint detector detects endpoints after an iterative process, while the applicants first detect a start position using a speech recognizer, which is then used to generate the plurality of pieces of speech data *(Appeal Brief, Page 10)*. In response, the examiner notes that claim 3 requires that "information of the start position of said speech region is provided by a speech recognition engine which performs the speech recognition," there is no mention of the order of processing. As such, Bi shows his endpointer as *part of* a speech recognition engine system that recognizes a user's voice *(voice recognition (VR) system, Fig. 1, Elements 10 and 22; and Col. 3, Lines 33-39)*. Thus, since the start position *(PRE_Start)* in Bi is *provided by* a speech recognition engine *(Fig. 1, Element 10)* because the endpointer is a part of this engine, this argument is not convincing.

With respect to dependent claims 4, 12, and 20, the appellants newly argue that the prior art of record fails to teach "that the information of the start position of the speech region is obtained by performing a recognition process on a first speech data by using the speech recognition engine, or is obtained by averaging speech data for several pieces of data from the start which would have been subjected to the recognition processing. As explained above, Bi's end-pointing process is a speech recognition process because it is performed by a speech recognition engine and is a process relating to speech recognition *(i.e., part of the processes performed by the speech recognizer) (Fig. 1, Elements 10 and 22; and Col. 4, Lines 37-57)*. Since the plurality of speech candidate periods are generated by Bi are all generated from by

shifting backwards from the *same* PRE_Start position *(i.e., continually adding additional data prior to the same initial PRE_Start index position)*, this PRE_Start position would be the average starting point for all backwards-extended region periods *(Figs. 2-3)*. Thus, the PRE_Start point is the *average start position as a result of the processing performed by the speech recognizer (Fig. 1, Element 10)*. As such, it appears that the applicant has mischaracterized the Office Action, and the arguments have been fully considered, but are not convincing.

With respect to dependent claims 5, 13, and 17, the appellants newly argue that the cited prior art does not in any way disclose that the start position of non-speech regions of a plurality of pieces of speech data are determined by changing the reading position in a speech buffer. In response, the examiner notes that Bi explicitly recites the use of a "data buffer" *(Col. 5, Lines 12-30)*, which has its reading index sequentially decremented in order to generate candidate speech periods *(see above)*. Thus, since Bi explicitly teaches the generation of multiple candidate speech periods by decrementing back from the start of a speech region *(PRE_Start located at the beginning of a strong speech region, See Fig. 3)* using a reading index of a "data buffer", this argument has been fully considered, but is not convincing.

(11) Related Proceeding(s) Appendix

No decision rendered by a court or the Board is identified by the examiner in the Related Appeals and Interferences section of this examiner's answer.

For the above reasons, it is believed that the rejections should be sustained.

Respectfully submitted,

/James S. Wozniak/

Patent Examiner, Art Unit 2626

Conferees:

/JW/,

/Patrick N. Edouard/

Supervisory Patent Examiner, Art Unit 2626

/Vijay B. Chawan/

Primary Examiner, Art Unit 2626

for Richemond Dorvil, SPE of Art Unit 2626